

# Trends And Services For Big Data Analysis On Cloud Data Centers

Varun Tiwari and Dr. Atul Dattatray Newase

Department of Application, Dr. A.P.J. Abdul Kalam University,  
Indore (M.P.) 452010, India

Corresponding Author Varun Tiwari

---

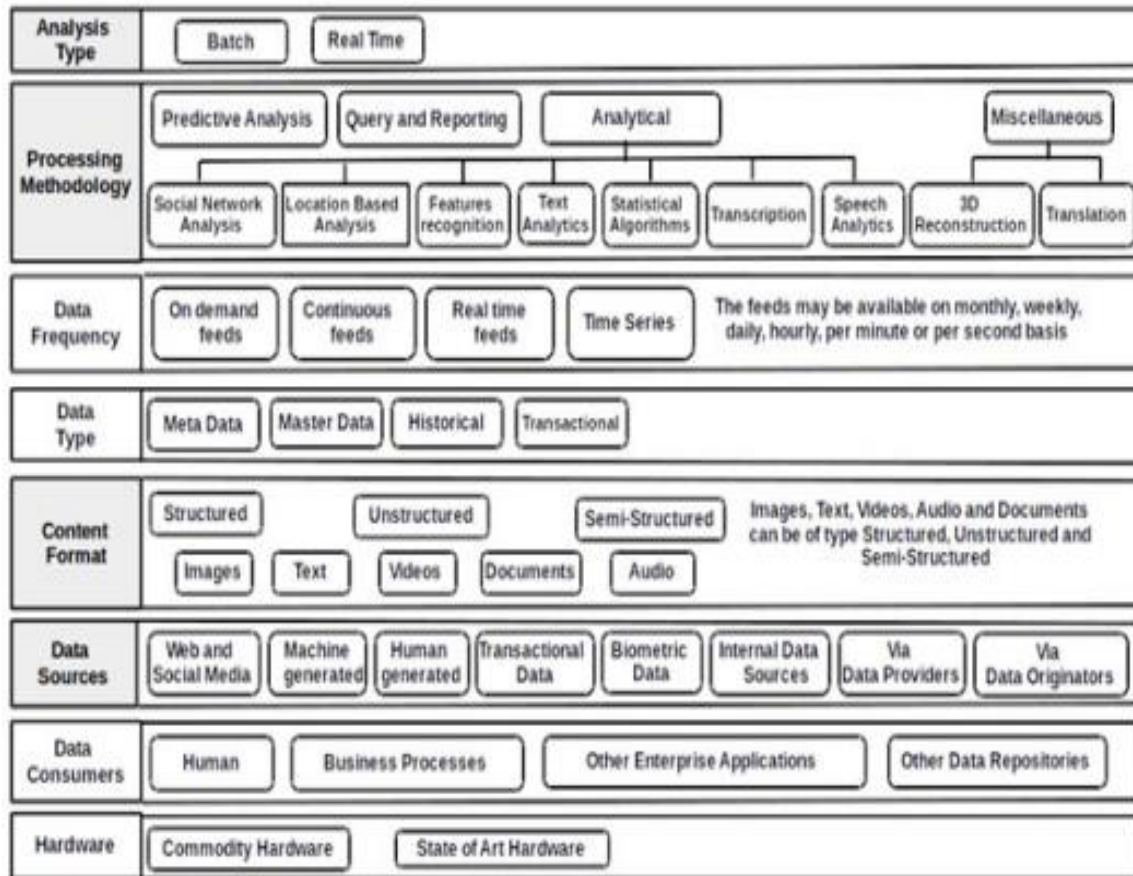
**Abstract:** This paper discusses about the current and future trends for big data on cloud data center, cloud data centers services, architecture of storing and processing big data in cloud data centers. Big data is a collection of large volume of data. Data centers is providing the facility to store, analyze, and process huge amount of data on networked computer. In today trends cloud computing are using for big data analytics. A cloud framework is used for developing, deploying and managing cloud-based applications on internet. So that cloud data centers are providing facility to run many big data applications over internet. In this research paper we also discuss about different types of cloud data centers. In this research paper we will discuss about past, present and future cloud data centers and challenges. How cloud data center will helpful for current trends and future directions of Big Data Analysis.

**Keywords:** Current and Future Trends, Cloud Computing, Cloud data centers Services, Cloud – Based Big Data Analytics, Cloud Data Center, Hadoop architecture

## I. INTRODUCTION

The term big data can come in different form, like non-structured and structure data such as financial, text, binary, multimedia files etc. Traditional database system widely used in Organization. You can easily manage and handle this structure data. But unstructured and semi structure data is requiring different techniques and tools to process and analyze. In cloud environment and parallel process architecture that enable parallelized data consume and investigation are the preferred approach to process such complex data. Structured Data, in this type of data, we have defined data types, formats and structures. All the data in well forms like data cubes, RDBMS, CSV files and spreadsheet. Semi-structured Data, XML Data files or self-description data files are called semi-structure data. Quasi-structured Data, the irregular data formats can be formatted with the need of some effort, tools and time is required to design this type of data. (For example, web clickable data which may contain incompatibility in data values and

formats). Unstructured data, data that does not have any underlying structure, in which text documents, PDF, images and videos can be included.[1]



**Figure1: Classification of Big Data**

Microsoft extended Douglas Laney’s 3Vs attributes to 6 Vs, which it added variability, veracity, and visibility:

1. **Volume** scale of data
2. **Velocity** analysis of flowing data
3. **Variety** different types of data
4. **Veracity** its emphases on reliability of data sources
5. **Variability** it indicates complexity of data set. It depends on that the number of variables in data sets
6. **Visibility** Complete visibility of data means full picture of data to make informative decisions.

Big data classification in different form like methodology, analysis, data type, data frequency, sources of data etc. Here we classify the data in the form of figure 1:

Cloud is a technology where many components are combined and shared resources through virtualization technology across the network. There are four components are available like:

1. Front-End (Client Infrastructure)
2. Back-End (Server, storage, services, application)
3. Cloud-Based Framework (Cloud Run Time Services)
4. Network

Cloud is working with the help of 4components. so, these four components are most important for running cloud application, data centers.cloud is provided three services for any organization:

1. **Software as a Service (SaaS):** This service provides the features for maintaining and delivering applications and software to organizations over the Internet. It provides a web interface for running application on different devices and OS over internet.
2. **Platform as a Service (PaaS):** It is middleware service. It provides the platform for creating and build application or service over on cloud. It also provides software deployment and configuration setting over on internet.
3. **Infrastructure as a Service (IaaS):** This service is a third-party provider. In this model don't need to purchase servers, networks or storage devices of any organization. Here third-party provider is providing to you all facility for manage software and applications. Only pay for the storage, server and network whenever you need.

#### **According to Cloud Computing Architecture There are four types of cloud:**

1. **Public cloud:** In this cloud service provider are used own computing resources. But these resources are shared and distributed many tenants over the internet. Due to sharing features operating cost and maintenance are less.
2. **Private cloud:** This cloud is owned and managed privately. Company provides own data center on own premises. It is more secure and customizable rather than public cloud.
3. **Hybrid cloud:** This is the combination of public and private cloud. This cloud contains Operating features of public cloud and data security capabilities of the private cloud. This cloud helpful for then organizations to migrate workloads between environments depending on their IT and data security requirements.
4. **Multi-Cloud:** This cloud uses multiple public cloud services. Multi cloud features provide greater flexibility to choose and deploy the cloud services that are most likely to satisfy varying organizational requirements.

## II. METHODOLOGY

In this research paper we using Systematic Literature Review Criteria for current trends and future directions of big data in cloud data center, Cloud data services, architecture of storing and processing big data on cloud data center, past, present and future cloud data centers. These SLR criteria is based on two phases like:

**(i) Planning:** Planning is divided in to two parts:

**(a) Need and Identification:**

The basic need of big data calculation on a cloud center is storage, memory, network, Virtual Machine etc. So, the first priority to check availability of these things on cloud data center. After that user phases lot of challenges on Cloud data center for management of Big Data. Check API Service, No SQL Service and SQL Services. After completing this task check future and current trends of this. Find an architecture for processing data on cloud like Hadoop.

**(b) Research Questions:**

(1) What are current and future trends for big data on cloud data center.

(2) How to work cloud data service?

(3) How to store and process large amount of data on Cloud Centers with architecture?

(4) Define past, present and future cloud data centers.

**(ii) Reporting:** At last, we provide conclusion which is based of this research paper.

## III. CLOUD DATA CENTERS SERVICES

**Data Centers:**Data center is most important and largest block of cloud. Cloud data centers is containing many computer nodes. The working of this computer nodes is to provide computation power, processing of data etc. It provides computation power for processing any resources which is acquired by the user.

This center contains many switches, which provide the feature to interconnect computer nodes and data center itself with outside world.

**Compute Nodes:** It is used for virtualization and compute power using by Virtual Machine. This virtual machine instances work on platform as a service level. Here IaaS have no role. Here VM share 1 or more CPU resources for allocation. According to resources availability PaaS increase memory, local storage and network quota according to condition and situation. The VM are available different types like Extra-large VM 8x and large 4x and small 2x. In this part CPU do relation 1 to 1 relation with virtual CPU (Physical machine shared by VMs of Multiple users). Computer node is used largest VM types. This Extra Large VM are used for Azure Cloud. When VM types are fully occupies physical node then no any other VM user will run on this physical machine. This data center provides many switches or link for sharing data among multiple users using VM.

**Deployment:** When a user hosts any application on virtual space this process is called deployment. Here VM assign a service on computer node for checking fault at the time of deployment. A load balancer balances all service/request at the time of processing data by VM. That reason User unknown this type working of VM in the data center due to virtualization. Deployment process is limited in size. This depends on constraints and priority of site. When you deploy large number of applications then need large number of CPU. Here we providing some examples related providers of cloud computing services like:

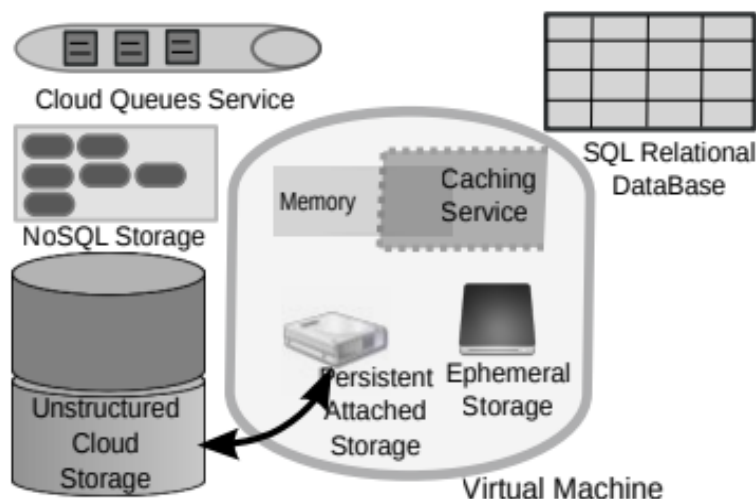
**1. Elastic Cloud Computing (EC2):** This service is used by Amazon. It's work for computing services and check IaaS level. Amazon is first vendors. It rented "de-facto" standards VM for cloud services to user. This VM provide the features to control over the software stack for user. This amazon cloud service offering to include data-oriented analytic platforms such as MapReduce for deployment.

**2. Microsoft Azure:** This is another cloud platform. It offers PaaS services and one IaaS. Here two roles are available Web roles and Worker roles. The VMs run either Windows or Linux-based OS. It used different software stack for deploy like java or .net.

**3. Google App Engine:** It is PaaS level cloud which is provided by Google. This app engine is provided facility to develop applications in several programming languages (e.g., Python, Java, PHP, go etc.) and integrate code extensions or technologies (e.g., Node.js, C++ etc.). It provides high networking performance and scaling for application deploy. It is a fine grain model. It charges VMs usage by the minute, after the first 10 minutes, unlike Amazon and Azure clouds which charge by the hour.

#### **IV. BIG DATA STORAGE ARCHITECTURE IN CLOUD**

Cloud data centers provide storage capacity and compute power for data management. It is used API services for getting and putting data for user. This API works by Virtual Machine. It is not free, its payable and prices depends on data size, time frame and operation on data. This data system contains lots of services which is shown in figure 2.



**Figure 2: Cloud Services for Big Data Storage**

**Ephemeral Storage:** In this phase it provides a virtual disk for storage large amount of data. The storage capacity depends on data size. (Azure used 2TB virtual disk for extra-large VM). However, the data stored in the virtual disks are ephemeral, as they persist only for the lifetime of the instance, being subject to losses caused by failures of the instance.

**Unstructured Cloud Storage:** In this data storage cloud offering for sharing application data. This data is available both to cloud compute nodes and to on-premise applications. It is distributed storage system.

**SQL Relational Databases:** It is primary option for storing application data. This relational database provides ACID guarantees (atomicity, consistency, isolation and durability). By using relational database cloud data center provide Database as a Service. It provides the facility to storage online relational databases and data management (retrieve, update and deploy etc.).

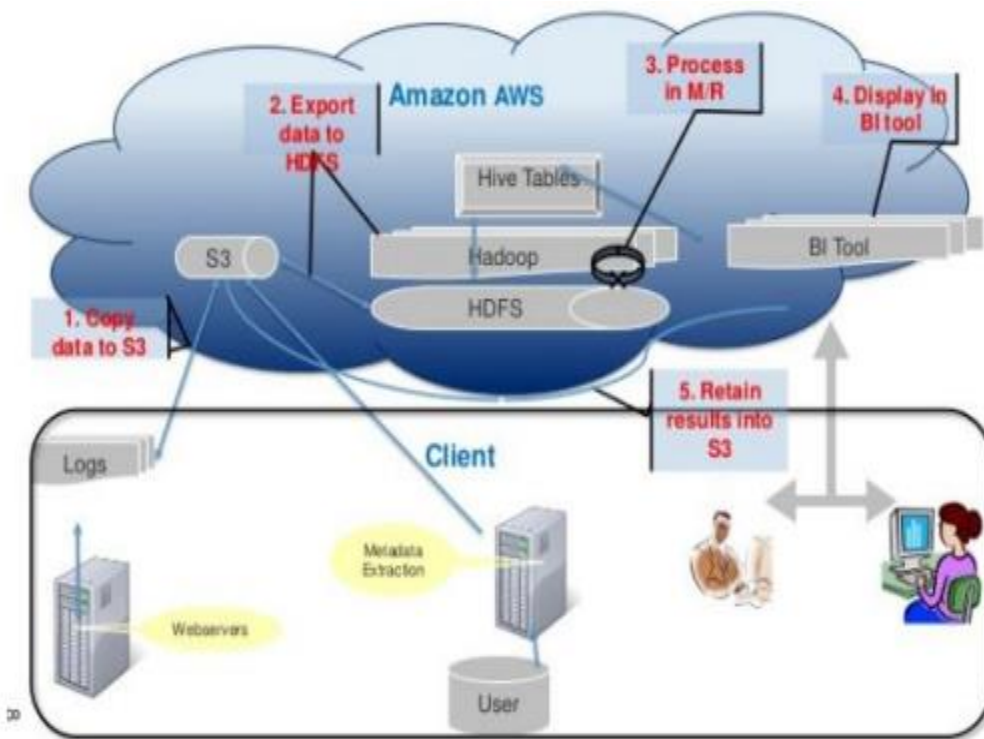
**NoSQL Relational Databases:** In this type data storage in the form of type-value or key-value pairs. There are lots of example of No-SQL like Cassandra, Google Bigtable, or Yahoo PNUTS.

**Cloud Queue Service:** It provide dedicated cloud for nodes synchronization. In this node can exchange small (e.g., 64 KB), short-term (e.g., 1 week) messages via the cloud queue service. It is designed as a communication for a high number of concurrent operations from many clients and to allow thousands of message transactions per second. Example are Azure Queue.

## **V. BIG DATA ARCHITECTURES (HADOOP) IN CLOUD DATA CENTER**

In this section we show Hadoop based cloud data center architecture. In this figure user collect data from different sources. After encrypt of these data. When data will be encrypted then data sent to storage on cloud data center. At the time of storing process, here it divides in cluster via HDFS. Here data will compute with the help of node. There are two types of nodes available like NameNode and DataNode. This NameNode is responsible for the data distribution to DataNode.

After complete HDFS task. Then data sent to starts for processing. For finding the result of these data uses hive or map reduce (API based Application as a Service). AWS is using Hadoop. This cloud service is used by Amazon. This amazon cloud service offering to include data-oriented analytic platforms such as MapReduce for deployment.



**Figure 3: Big Data Architecture (Hadoop) in Cloud Data Center**

## VI. CURRENT AND FUTURE TRENDS FOR BIG DATA ON CLOUD CENTER

Cloud is most important part of IT industry. There are lot of big data tools and cloud technology available like R cloud, Aws, Azure, Google Cloud etc. The existing system migrated with cloud the providers. Today many companies investing in the field of cloud computing. Currently we are using public, private and hybrid cloud. In future the most important changes on cloud will become called multi-Cloud. This type of cloud technique will change the future of IT Industry and image of data centers. There are many key features are available for current trends for big data on cloud:

1. Many new functions are evolving for big data in cloud world.
2. Cloud is a key for many emerging technologies on Big Data Field
3. BI introduces in Current Trends.
4. Currently hybrid cloud is growing fastly.

5. Visualization model will include on big data cloud.
6. AI, ML include for Big Data Analysis and prediction on Cloud like Google Colab, Jupiter and R cloud works SaaS and PaaS form.

In future model many things are include like edge computing, IoT devices etc. In this part we are showing future trends. There are many key features are available for future trends for Big Data in Cloud:

1. Big Data provide stronger relation on Cloud Storage (streaming data, observation data perform easily from cloud).
2. In future Data Fabric Technology will develop in the cloud. Because DFT is used AI and ML automation for big data.
3. Multi Cloud will helpful for Big Data Cloud Center. It will used for distribution workload across multiple cloud environment. The benefit of this model that you can store all sensitive information on private cloud separately.
4. In future cloud is integrating security with ML and AI. It will provide to facility to automatic threat detection and prevention using cloud data centers.
5. In future Edge computing will used over Cloud Data Center. This edge computing technology will manage massive data generated by the IoT devices.
6. In future cloud computing will leads to massive increase data storage capacity.
7. In future many companies will adopt a serverless platform.

## **VII. CHALLENGES FOR BIG DATA ANALYTICS ON CLOUD CENTER**

The main challenge is handling numerous amounts of data for storing, processing, analyzing and prediction the large amount of set of information on various cloud data centers. There are many key challenges for Big Data Analysis on Cloud mode.

1. Main problem is knowledge about the modern technologies.
2. Understanding to find the massive data according to their characteristics
3. Data is growing rapidly. The main challenge is storage capacity on Data Centers.
4. Better tools selection for Big Data Analysis on Cloud.
5. Data Security on Cloud is big challenge.
6. Challenges for big data analytics and deep learning on cloud need more computation (need faster GPU used for interconnection and faster CPU for memory). Need more fast networking technology.
7. The main need of Software-Defined storage in future.



## VIII. CONCLUSION

This research is helpful for solve the problem of Big Data storage, processing, analysis and prediction data on cloud. We are using cloud data centers to solve that problem. There are many cloud data services available and many cloud data storage platform which is provide the facility for analysis of data on cloud. There are many key challenges for Big Data Analysis on Cloud. In this research paper we introducing various current trends and future trends for Big Data Analysis on Cloud Data Centers. In these days big data growing faster. In future will need more storage, security, multi cloud platform for running complex application in distributed mode.

## REFERENCES

1. Manoj Muniswamaiah, Tilak Agerwala and Charles Tappert Seidenberg “Big Data in Cloud Computing Review and Opportunities “Volume 7 issue 1 (ISSN:2320-0294). International Journal of Computer Science & Information Technology (IJCSIT) Vol 11, No 4, August 2019 DOI: 10.5121 (School of CSIS, Pace University, White Plains, New York).
2. G. Ramakrishna CSE Department “An Empirical Study to Design a Group Communication Framework for Securing BIG DATA in HADOOP Based Cloud Data Center” Volume 7, Issue 4, April 2017 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.
3. Varun Tiwari, Atul D. Newase “Bca Students Result Analysis Using R” Volume 20, No 8 (2022), NeuroQuantology.
4. Varun Tiwari, Atul D. Newase "A Study on Big Data Analytical Tools" in telematique Vol 21, No 1 (2022).
5. Varun Tiwari, Vikas Rao Vadi and Shafiqul Abidin “Big Data Analysis using R and Hadoop” in IEMIS 2020 2nd International Conference in AISC Series Springer.
6. Varun Tiwari, Vikas Rao Vadi and Mukta Sharma “WOT Enabled Smart Cities integrated Big Data Analytics”, IJCA.
7. Varun Tiwari, Atul D. Newase and Vikas Rao Vadi “Design and Implement ICFR Analyzer Model for Data Analysis of COVID-19 in India” in Design Engineering (Toronto) ISSN: 0011-9342.
8. Varun Tiwari, Atul D. Newase, Vikas Rao Vadi and Shafiqul Abidin “Skytree Contribution in Mining Massive Datasets of Society 4.0” in Gradiva Review Journal ISSN NO: 0363-8057.
9. Varun Tiwari, Kusum Lata Bharti and Tanvir A. Abbasi “A Brief Survey on Data Mining using Latest Technology (Big Data)”, in IJRASET (ISSN:2321-9653).
10. Varun Tiwari, Shafiqul Abidin and Rajeev Kumar “Detecting Malware by Data Mining”, in IJESM
11. Kusum Lata Bharti, Tanvir A. Abbasi and Varun Tiwari “A Brief Survey on Data Mining

- using Latest Technology (Big Data)”, in IJRASET (July 2018) (ISSN:2321-9653).
- 12.** Varun Tiwari, Shafiqul Abidin, Rajeev Kumar “Detecting Malware by Data Mining”, in IJESM Volume 7 issue 1 (January 2018) (ISSN:2320-0294).